Accounting for Pairwise Heterogeneity in Bilateral Trade Flows: A Stochastic Varying Coefficient Gravity Model

Vangelis Tzouvelekas

(*Dept. of Economics, University of Crete, Greece*)

Please direct correspondence to:

Vangelis Tzouvelekas

Dept. of Economics, University of Crete

University Campus, 74100, Rethymno, Crete, Greece.

Tel.: +30-831-77426; Fax: +30-831-77406;

E-mail: vangelis@econ.soc.uoc.gr or vangelis@cha.forthnet.gr

Accounting for Pairwise Heterogeneity in Bilateral Trade Flows: A Stochastic Varying Coefficient Gravity Model

Abstract

This paper suggests an alternative way for estimating the gravity equation that takes into consideration country-pair heterogeneity in bilateral trade flows. Specifically, a stochastic varying coefficient gravity model based on Hildreth and Houck's (1968) random coefficient regression is proposed, that eliminates heterogeneity bias inherent in standard econometric methods. The results indicate that the standard gravity estimates can differ substantially from what is obtained when heterogeneity is accounted for.

1. Introduction

During the last thirty years the gravity equation has been extensively applied in empirical studies of international economics to provide valuable insights into the functioning of both interregional and international bilateral trade flows. Specifically, the gravity equation has been used to analyze the impact of a variety of policy issues such as trading groups, political blocks, patent rights and various trade distortions. Despite of it's extensive use, the early applications of the gravity model had lacked rigorous theoretical underpinnings and was long criticized for being rather *ad hoc*.

Recently Deardorff (1998) and Baier and Bergstrand (2001) have shown that the gravity equation of bilateral trade can be derived from a factor endownment model consistent with classical theories of trade. This was an extension of the earlier works by Anderson (1979), Helpman and Krugman (1985), Bergstrand (1985; 1989) and Hummels and Levinsohn (1995) who also derived theoretically the gravity equation as a reduced form from a general equlibrium model of international trade. Although

the theoretical foundations seems that have been established, the empirical estimation of the gravity model has not been addressed in detail.

In it's simplest form the gravity equation assumes that the amount of trade between two countries or regions is increasing in size, as measured by national income, and decrease in the cost of transportation between them, as measured by the bilateral distance between countries.[1] In other words, exporter and importer GDPs can be interpreted as the production and absorption capacities of the exporting and importing countries, respectively.

Linnemann (1966) suggested an augmented form of the gravity equation including population or per capita income as an alternative measure of country's size. As noted by Bergstrand (1989), regardless the particular specification of the gravity model, the prime purpose of any empirical application is to allow for non-homothetic preferences in the importing country and to proxy for the capital-labor ratio in the exporting country.

Through years other variables reflecting factors that make trade easier or more difficult, such as political considerations, preferential and free trade agreements, customs unions, tariff levels and neighborhood have been taken into account in gravity model specification. However, not all the factors affecting bilateral trade flows are readily observable. Variables such as historical links, cultural similarities etc., though they affect significantly bilateral trade flows, are usually unobservable and thus difficult to quantify empirically. For instance, if US consumers due to cultural reasons have a stronger preference for British made goods over German made goods, then if all else are equal, the US will import more from UK than from Germany.

Hence, exogenous factors not readily observable introduce a country-pair heterogeneity which should be taken into account in the econometric estimation of the gravity equation. With such heterogeneity a country may trade different amounts from two other countries even though the two markets have the same GDP and are equidistant from each other. This may happen due to historical, cultural, political or geographical factors that affect the level of bilateral trade flows and are correlated

---

[1] As noted by many authors the costs of distance may extend well beyond freight charges (i.e., Anderson and Marcouiller, 1999). Thus, while distance has always been a key variable in gravity models, authors have never been sure exactly what costs distance represents.

with the independent variables included in the gravity equation. In this case the standard econometric estimation (i.e. OLS, MLE) of the gravity equation, it is more likely to provide biased estimates.

Several attempts have been made in the literature to control for this heterogeneity bias including variables such as whether trading partners share a common language, have had a colonial history, are in military alliance etc. Nevertheless, these exogenous factors are often difficult to quantify empirically. In order to overcome the problem several authors have introduced dummy variables as an approach to include in the model specification these exogenous factors. However, dummy variables are simplistic in nature and thus they do not account for real trade diversion effects. It is quite important therefore, to find alternative ways to account for the impact of these factors in bilateral trade flows in the empirical estimation of the gravity equation.

In a recent paper Cheng and Wall (1999) found that in the presence of country-pair heterogeneity the standard estimation methods tend to underestimate trade between high-volume traders and overestimate it between low-volume traders. They suggest the use of standard panel data estimators to allow for the intercepts of the gravity equation to be specific to each trading pair in order to account for pairwise heterogeneity. Bayoumi and Eichengreen (1997) and Mátyás (1997) have also proposed models to handle pairwise heterogeneity each of which can be expressed as a restricted version of the fixed effects panel data model.

Besides the need of panel data, in this case the impact of the unobservable variables on bilateral trade flows is assumed to be neutral with respect to each trading pair. The estimated parameters of the gravity equation are assumed to be the same across country-pairs except of the intercept term. Nevertheless, it is logically to assume that the impact of the explanatory variables included in the gravity equation is not affecting uniformly bilateral trade flows among country-pairs. The diversity of pairwise trade behavior, would lead to parameter variation across country pairs. In such cases the constant slope but varying intercept coefficients do not appear to be meaningful.

The objective of this paper is to present an alternative way for estimating the gravity equation that accounts for country-pair heterogeneity. A stochastic varying coefficient gravity model is suggested that allows for modeling the heterogeneity in

the functional relationship between bilateral trade flows and explanatory variables. The model is based on Hildreth and Houck (1968) random coefficient regression popularized by Swamy (1970). The empirical application is based on cross-section data of bilateral trade flows between EU member states.

The rest of the paper is organized as follows: next, drawing from the relevant literature, the theoretical underpinnings of the conceptual model used are discussed; next, data and the empirical results along with their implications derived from the analysis are interpreted in the penultimate section; concluding remarks are summarized in the final section.

2. Methodological Framework and Estimation Procedure

The basic log-linear[2] formulation of the gravity model under market equilibrium conditions of demand and supply systems can be derived as follows:

$$ln\,Y_i = \beta_0 + \beta_1\,ln\,G_i^X + \beta_2\,ln\,G_i^M + \beta_3\,ln\,N_i^X + \beta_4\,ln\,N_i^M + \beta_5\,ln\,D_i + v_i \quad (1)$$

where, Y_i is the current value of bilateral trade flows; $G_i^{X(M)}$ is the nominal gross domestic product (GDP) of exporting (X) and importing (M) countries; $N_i^{X(M)}$ is the population in exporting and importing country; D_i is the distance between countries economic centers that serves as a proxy of transportation and information cost; v_i is the usual *iid* error term and; *i=1, 2, …, N* are the country-pairs.

In the above formulation both the intercept term and the slope coefficients are common to every country-pair in the sample. In other words, the impact of the explanatory variables on bilateral trade flows is the same among country-pairs. However, as explained at the outset the response of GDP, population or distances is strongly influenced by factors that are not readily observable. Hence, it would be desirably to incorporate their influences into the model. If these factors vary among country-pairs but are unobserved, regarding *β's* as the means of a random response may be better than assuming that these are constant.[3]

---

[2] Recently, Sanso, Guairan and Sanz (1993) using the Box-Cox transformation function concluded that the log-linear form, while not optimal, is a fair and ready approximation of the optimal form.

[3] This may be an oversimplification of the behaviour of beta coefficients if *β's* vary in a systematic way with these quantifiable variables. However, given that such information is not always available, the present assumption seems reasonable. In either case this assumption can be verified using the

In this case the gravity model in (*1*) can be written with the following general form:

$$ln\ Y_i = \beta_{0i} + \sum_{k=1}^{K} \beta_{ki}\ ln\ X_{ki} + v_i \qquad (2)$$

where, $\beta_{ki} = \bar{\beta}_k + u_{ki}$, $\bar{\beta}_k$ is the mean response of bilateral trade flows to a unit change in the *kth* independent variable and β_{ki} is the actual response rate in every country-pair; $E(\beta_{ki}) = \bar{\beta}_k$, $E(u_{ki}) = 0$; $var(u_{ki}) = \sigma^2_{uki}$; $cov(u_{ki}, u_{li}) = 0$ for $k \neq l$; $\forall k,l = 0,...,K$ and; *X* is the vector of independent variables ($Y^{X(M)}$, $N^{X(M)}$, *D*). Under the above assumptions the model in (*2*) can be written as:

$$ln\ Y_i = \bar{\beta}_{0i} + \sum_{k=1}^{K} \bar{\beta}_{ki}\ ln\ X_{ki} + \varepsilon_i + v_i \qquad (3)$$

where, $\varepsilon_i = u_{0i} + \sum_{k=2}^{K} u_{ki}\ ln\ X_{ki}$; $E(\varepsilon_i) = 0$; $var(\varepsilon_i) = \sigma^2_{u0i} + \sum_{k=2}^{K} \sigma^2_{uki}\ ln^2\ X_{ki}$; $cov(\varepsilon_{ki}, \varepsilon_{li}) = 0$ for $k \neq l$; $E(\varepsilon_i, v_i) = 0$ and; $var(v_i) = \sigma^2_{vi}$.

The model in (*3*) is essentially Hildreth and Houck's (1968) random coefficient regression model and can be consistently estimated using generalized least squares (GLS). The variance-covariance matrix for the composite disturbance term is block-diagonal with the *ith* diagonal block given by:

$$\Sigma_i = X_i \Delta X_i' + \sigma^2_{uoi} \qquad (4)$$

where Δ is the matrix of the variances of u_{ki}. Under Swamy's (1970) assumptions the best linear unbiased estimator of *β's* is obtained:

$$\hat{\beta} = \left(X'\Sigma^{-1}X\right)^{-1}\left(X'\Sigma^{-1}Y\right) \qquad (5)$$

with a covariance matrix given by:

$$var(\hat{\beta}) = \left(X'\Sigma^{-1}X\right)^{-1} \qquad (6)$$

To make the estimator operational, however, we need to estimate the variances of the random coefficients. For doing so we use the OLS residuals given by:

$$e = Y - Xb = M \cdot Y = M \cdot \varepsilon \qquad (7)$$

Breusch and Pagan (1979) LM test.

where $M = I - X(X'X)^{-1}X'$ is the idempotent matrix of order n and b is the OLS estimator of β. The *ith* element of e equals $e_i = \sum_j^n m_{ij}\varepsilon_j$, where m_{ij} is the corresponding element of the idempotent matrix. Hence, it holds that $Ee_i = 0$ and $Var(e_i) = \sum_j^n m_{ij}^2 \sigma_{jj} = E(e_i^2)$ where σ_{jj} is the corresponding diagonal element of Σ. We may write n observation compactly as:

$$E\overset{\bullet}{e} = \overset{\bullet}{M}\overset{\bullet}{X}\overset{\bullet}{\sigma} \tag{8}$$

where the dot indicates the squared elements of the corresponding matrices. For instance $\overset{\bullet}{X} = X \otimes X$ where $\otimes$ represents the Kronecker matrix product. The regression relation conditional on the explanatory variables implied by (*8*) can be written as:

$$\overset{\bullet}{e} = \overset{\bullet}{M}\overset{\bullet}{X}\overset{\bullet}{\sigma} + \omega = Z\overset{\bullet}{\sigma} + \omega \tag{9}$$

where ω is an *iid nx1* disturbance vector. By applying OLS to (*9*) we can obtain an estimate of the vector of variances of the random coefficients, σ:

$$\hat{\sigma} = (Z'Z)^{-1}Z'\overset{\bullet}{e} \tag{10}$$

By replacing them into (*4*) we obtain an estimate of the variance-covariance matrix of the composite error term, $\hat{\Sigma}$, and the GLS estimator in (*5*) becomes operational. Individual response coefficients can be then obtained using the approach put forward by Griffiths (1972) as follows:

$$\hat{\beta}_{ki} = \hat{\beta}_k + \phi X'[X\phi X']^{-1}(Y - X\hat{\beta}_\kappa) \tag{11}$$

where $\phi = diag(\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_k)$ is obtained from the OLS estimation of (*9*). Estimators of mean and individual parameter vectors are best linear unbiased in a sense that where the expectation operator is an unconditional one (Griffiths, 1972).

The assumption of producer specific coefficients can be examined using the standard Breusch-Pagan LaGrange multiplier test for the class of heteroscedastic error models (Breusch and Pagan, 1979). The LM test has the same asymptotic properties as the likelihood ratio test but it is computationally much simpler.

3. Data and Empirical Results

The empirical analysis was carried out for the 1997 period and the countries considered were the 15 EU countries. Data on the volume of trade, population and GDP were taken from *Eurostat*. The distances between the capitals were obtained using an Internet service which takes data from the US Census and a supplementary list of cities around the world.

Generalized least squares estimates[4] of the stochastic varying coefficient gravity model are presented in Table 1. For comparison purposes the OLS estimates assuming common intercepts and slope coefficients are also presented in the same table. The difference between the two models is glaring and statistically significant. The Breusch-Pagan LM-test accept the hypothesis of random coefficient variation lending support for the stochastic varying coefficient gravity model specification. Calculated chi-square statistic is 86.12, well above the corresponding tabulated value at the 1% level of significance. Hence, country-pair heterogeneity is important factor in explaining bilateral trade flows among EU countries.

Mean response coefficients $\left(\bar{\beta}_{0i}, \bar{\beta}_{ki}\right)$ of the stochastic varying coefficient gravity model along with their corresponding standard errors are presented in the fourth and fifth columns of Table 1. According to these estimates a 10% increase in a country's GDP would results *ceteris paribus* a 4.04% rise in exports and a 3.59% rise in imports on the average. Exports seems to be labor intensive and income inelastic, as indicated by the positive sign of origin and destination population, respectively. Finally, transportation and information cost are affecting negatively the volume of trade between EU countries.

As it is clearly shown in both Table 1 (last three columns) and Figure 1, the variation between trading-pairs of the average parameter estimates is substantial. This in turn implies that individual contributions of independent variables to bilateral trade flows vary significantly across country-pairs due to the existence of pairwise heterogeneity. The highest variation is observed for importing country's population and the lowest for the distance between economic centers. Specifically, the parameter estimate for exporting country's GDP ranges between 0.146 and 0.505, for importing

---

[4] The estimation of the stochastic varying coefficient gravity model was carried out using a simple and implementable program in Gauss (Ver. 3.2.26).

country's GDP between 0.142 and 0.428, for exporting country's population between 0.033 and 0.352, for importing country's population between 0.803 and 1.319 and for distances between -0.268 and -0.163.

The econometric estimates of the OLS and SVCG models are different in terms of both the magnitude and the sign of the respective coefficients. The parameter estimate of exporting country's GDP is negative and statistically significant implying that an 10% increase in GDP would result *ceteris paribus* a decrease in the volume of exports. On the other hand, the OLS estimates of the gravity model clearly indicates that the absolute value of all the parameter estimates except that of the distances between economic centers is increased. The difference is more in evident for exporting country's population and importing country's GDP.

It is noteworthy the fact that according to the OLS estimation of the gravity equation, exporting country's population and importing country's GDP are the foremost important factors determining bilateral trade flows. However, the SVCG model parameter estimates suggest that importing country's population and exporting country's GDP are mainly determining bilateral trade flows, whereas exporting country's population has the lowest impact in country's trading behavior.

These findings is rather important indicating that the impact of the explanatory variables included in the gravity equation is misinterpreted if the pairwise heterogeneity bias is not accounted for during the econometric estimation of the model. It is obvious from the results that restricting the country-pair effects to zero, as does the OLS estimation of the gravity equation, has significant effect on the results. Hence, in the absence of any economic *a priori* arguments for believing that the parameters of the gravity equation are the same across trading partners, the stochastic varying coefficient gravity model seems to be a more appropriate specification.

4. Conclusions

During the last decades the gravity equation has been used extensively in analyzing bilateral trade flows in both developed and developing countries. However, despite of the recent flurry of empirical work aimed to establish the theoretical foundations of the gravity equation, it's empirical estimation has not been addressed in detail. Specifically, not all the factors affecting trading behaviour among countries

are readily observable and thus quantifiable. This ommission introduce however, a country-pair heterogeneity in bilateral trade flows that the standard econometric methods do not take into consideration in the estimation of the gravity equation.

The purpose of this paper is to put forward an alternative way for the estimation of the gravity equation that accounts for the existence of differences in the trading behaviour among partners. Specifically we suggest a stochastic varying coefficient gravity model (SVCG) that allows for modeling the heterogeneity in the functional relationship between bilateral trade flows and explanatory variables. The model is essentially based on Hildreth and Houck's (1968) random coefficient regression popularized by Swamy (1970). The empirical application is based on cross-section data of bilateral trade flows between the EU member states for the 1997 period.

The empirical results reveal that the OLS estimation of the gravity equation suffer from heterogeneity bias due to omitted or misspecified variables. The econometric estimates of the OLS and the SVCG models are different in both the signs and the magnitudes of the respective coefficients. Hence, country-pair heterogeneity is an important factor that should be taken into consideration in the econometric estimation of the gravity equation.

References

Anderson , J.E. and D. Marcouiller, "Trade, Location and Security: An Empirical Investigation," *NBER Working Paper* No. 7000 (1999).

Anderson, James E., "A Theoretical Foundation for the Gravity Equation," *American Economic Review* 69 (1979), 106-116.

Baier, Scott L. and Jeffrey H. Bergstrand, "The Growth of World Trade: Tariffs, Transport Costs, and Income Similarity," *Journal of International Economics* 53 (2001), 1-27.

Bayoumi, T. and B. Eichengreen, "Is Regionalism Simply a Diversion? Evidence from the Evolution of the EC and EFTA," in T. Ito and A.O. Krueger (eds.), Regionalism versus Multilateral Trade Arrangements (University of Chicago Press, 1997).

Bergstrand, Jeffrey H., "The Generalized Gravity Equation, Monopolistic Competition and the Factor Proportions Theory in International Trade," *Review of Economics and Statistics* 71 (1989), 143-153.

Bergstrand, Jeffrey H., "The Gravity Equation in International Trade: Some Microeconomic Foundations and Empirical Evidence," *Review of Economics and Statistics* 67 (1985), 474-481.

Breusch, T.S. and A.R. Pagan, "A Simple Test for Heteroscedasticity and Random

Coefficient Variation," *Econometrica*, 47 (1979), 1287-1294.

Cheng, I-Hui, and Howard J. Wall, "Controlling for Heterogeneity in Gravity Models of Trade," The Federal Reserve Bank of St. Louis Working Paper Series No 99-010A, 1999.

Deardorff, Allan V., "Determinants of Bilateral Trade: Does Gravity Work in Neoclassical World?," in J.A. Frankel (ed.), *The Regionalization of the World Economy*, (University of Chicago Press, 1998).

Griffiths, William E., "Estimation of Actual Response Coefficients in the Hildreth-Houck Random Coefficient Model," *Journal of the American Statistical Association* 67 (1972), 633-635.

Helpman, Elhanan, and Paul Krugman, *Market Structure and Foreign Trade* (Cambridge, MA: MIT Press, 1985).

Hildreth, Clifford, and James P. Houck, "Some Estimators for Linear Model with Random Coefficients," *Journal of American Statistical Association* 63 (1968), 584-595.

Hummels, D. and J. Levinshon, "Monopolistic Competition and International Trade: Reconsidering the Evidence," *Quarterly Journal of Economics* 110 (1995), 799-836.

Linnemann, Hans, *An Econometric Study of International Trade Flows* (Amsterdam: North-Holland, 1966).

Mátyás, László, "Proper Econometric Specification of the Gravity Model," *The World Economy* 20 (1997), 363-368.

Sanso, Marcos, Rogelio Guairan, and Fernando Sanz, "Bilateral Trade Flows, the Gravity Equation, and Functional Form," *Review of Economics and Statistics* 75 (1993), 266-275.

Swamy, P.A.V.B., "Efficient Inference in Random Coefficient Regression Model," *Econometrica* 38 (1970), 311-323.

Table 1. Parameter Estimates of the Stochastic Varying Coefficient Gravity Model

Variable	OLS		Stochastic Varying Coefficient Gravity Model				
	Estimate	StdError	Estimate	StdError	Max	Min	StdDev
Constant (β_0)	-1.035	(0.215)*	-0.737	(0.126)*	-0.687	-0.776	0.015
GDP_i (β_1)	-1.283	(0.617)*	0.404	(0.225)**	0.505	0.146	0.062
GDP_j (β_2)	1.884	(0.564)*	0.359	(0.168)**	0.428	0.142	0.049
N_i (β_3)	2.328	(0.767)*	0.129	(0.082)**	0.352	0.033	0.057
N_j (β_4)	1.375	(0.556)*	1.178	(0.404)*	1.319	0.803	0.084
D_{ij} (β_5)	-0.478	(0.277)*	-0.195	(0.095)**	-0.163	-0.268	0.016
Breusch-Pagan LM test:			66.12*				

* significant at the 1% level; ** significant at the 5% level.

Figure 1. Variation of Parameter Estimates